



> Semantic Web Use Cases and Case Studies

Case Study: Use of Semantic Web Technologies on the BBC Web Sites

Yves Raimond, BBC, United Kingdom

January 2010



General Description

The BBC is the largest broadcasting corporation in the world. Central to its mission is to enrich people's lives with programmes that inform, educate and entertain. It is a public service broadcaster, established by a Royal Charter and funded, in part, by the licence fee that is paid by UK households. The BBC uses the income from the licence fee to provide public services including 8 national TV channels plus regional programming, 10 national radio stations, 40 local radio stations and an extensive website, bbc.co.uk.

Linking microsites for cross-domain navigation

The BBC publishes large amounts of content online, as text, audio and video. Historically the website has focused largely on supporting broadcast brands (e.g. Top Gear) and a series of domain-specific sites (e.g. news, food, gardening, etc.). That is, the focus has been on providing separate, standalone HTML sites designed to be accessed with a desktop Web browser. These sites can be very successful, but tend not to link together, and so are less useful when people have interests that span programme brands or domains. For example, we can tell you who presents Top Gear, but not what else those people have presented. As a user it is very difficult to find everything the BBC has published about any given subject, nor can you easily navigate across BBC domains following a particular semantic thread. For example, until recently you weren't able to navigate from a page about a programme to a page about each artist played in that programme.

This lack of cross linking has also limited the type of user interaction the BBC is able to offer, for example, it is a complex piece of work to recontextualise content designed for one purpose (e.g. a programme web site) for another purpose or to extract the underlying data and visualize it in a new or different way. This is because of a lack of integration at a data level making it difficult to repurpose and represent data within a different context.

When building hand-crafted programme web sites, it also means that only some programmes could be covered. As the BBC broadcasts between 1,000 and 1,500 programmes a day, the long tail of our programming didn't get any web presence. Hand-crafted web sites are also harder to maintain. They often get forgotten and left unmaintained, or even removed.

Making data available to developers

The BBC, through [Backstage](#), has made 'feeds' available for third party developers to build, non-commercial, mash-ups. However, these feeds suffer from the same or similar issues to the microsites namely they lack interlinking. That is, it is possible to get a feed of latest news stories but it's not easy to segment that data into news stories about 'Lions'. Nor is it possible to query the data to extract the specific data required.

The Solution

A web presence for all BBC programmes

[BBC Programmes](#) was launched in Summer 2007. Its goal is to provide a web identifier, with associated HTML pages and machine-readable feeds (RDF/XML, JSON and XML), for every programme the BBC broadcasts—allowing other teams within the BBC to incorporate those pages into new and existing programme support sites,

TV Channel and Radio Station sites, and cross programme genre sites such as food, music and natural history. For example, BBC Programmes publishes the following page about the “Life” brand:

The screenshot shows the BBC Life website interface. At the top, there is a navigation bar with the BBC logo, a search bar, and a link to 'Explore the BBC'. Below this is a large 'Life' logo and a BBC One logo. The main content area is divided into several sections:

- Life**: A featured article about 'Creatures of the Deep' by David Attenborough, featuring a red frog on a green leaf. It includes a 'read more' link and links for 'PROGRAMMES: on BBC iPlayer (8) | clips (5) | coming up (8)' and 'PREVIOUS PROGRAMMES: by year (9)'.
- Coming up**: A section titled 'TODAY, 22:40 on BBC One' for 'Creatures of the Deep' on 8/10, with a description: 'A look at marine invertebrates, including Humboldt squid, starfish and giant octopi.' It also mentions '7 more programmes coming up in the next 7 days (including repeats)'.
- Available now on BBC iPlayer**: A large video player for 'Creatures of the Deep' with a 'CLICK TO PLAY' button and a 'WATCH IN POP-OUT PLAYER' button. The video player shows a jellyfish underwater.
- Links**: A list of links including 'Wildlife Finder: explore the BBC nature archive and find out more about the animals featured', 'Earth Explorers: follow the latest Life expeditions', and 'Filming Techniques: stories from behind the scenes (pdf)'.
- Learn more**: A section featuring 'The Open University' logo and links to 'Free Tree of Life poster and interactive', 'Play the Tree of Life game', and 'Watch Life video extras'.
- Watch more stunning videos**: A section titled 'WILDLIFE FINDER' with links to 'Watch David Attenborough's favourite moments', 'Wildlife stories from programme makers on location', and 'Earth News, reporting life on earth'.

At the bottom of the video player, there is a progress bar showing '0:00.00 / 1:00.00' and a 'Watch now (59 minutes)' button. Below the video player, there is a 'MORE:' section with links for 'programme information | related links | clips | chapters' and a note: 'Available since today with 21 days left. 8/10. A look at marine invertebrates, including Humboldt squid, starfish and giant octopi.'

And the following page about an episode of the “Life” brand:

The screenshot displays the BBC Life website interface. At the top, there is a navigation bar with the BBC logo, a search bar, and a 'Text only | Help' link. The main header features the 'Life' logo and the 'BBC one' logo. The central content area is titled 'Creatures of the Deep' and contains a video player. The video player shows a large jellyfish in a dark blue underwater environment. A 'CLICK TO PLAY' button is overlaid on the video. Below the video player, there is a progress bar showing '0:00.00 / 1:00.00' and a 'WATCH IN POP-OUT PLAYER' button. A chapter list below the video shows '7 Chapters within this programme:' with buttons for chapters 1 through 7. A blue banner below the chapters states 'This programme has been divided into chapters' and 'Start playing chapter 1'. To the right of the video player, there is a 'More details' section. It indicates 'EPISODE 8 OF 10 FROM' and lists 'Life' as the series. Under 'IN THIS SERIES', it shows 'Previous: < Hunters and Hunted' and 'Next: Plants >'. A '21 days left to watch' notification is present. The 'DURATION' is listed as '59 minutes'. Below this is a 'Learn more' section featuring 'The Open University' logo and three links: 'Free Tree of Life poster and interactive', 'Play the Tree of Life game', and 'Watch Life video extras'. At the bottom right, there is a 'Watch more stunning videos' section with a 'WILDLIFE FINDER' video thumbnail and three links: 'Watch David Attenborough's favourite moments', 'Wildlife stories from programme makers on location', and 'Earth News, reporting life on earth'.

The Web as a Content Management System

[BBC Music](#) follows the same principles as BBC Programmes, and provides a web identifier for every artist the BBC has an interest in (featured in music programmes, in BBC events, etc.). BBC Music is underpinned by the Musicbrainz music database and Wikipedia, thereby linking out into the Web as well as improving links within the BBC site. BBC Music takes the approach that the Web itself is its content management system. Our editors directly contribute to Musicbrainz and Wikipedia, and BBC Music will show an aggregated view of this information, put in a BBC context.

For example, BBC Music publishes the following page about “Bat for Lashes”:

BBC Text only Help Search Explore the BBC

Music BETA QUICK FIND Enter an artist name ...

BBC Music > Artists > Bat for Lashes

Bat for Lashes

PLAYED MOST ON **BBC RADIO** **6music**

Born 25 October 1979.
Bat for Lashes is a performance name for the person [Natasha Khan](#).



Now On The BBC

Bat For Lashes on Later...
Watch Bat For Lashes performing Sleep Alone on Later With Jools Holland.



Biography

Natasha Khan (born 25 October 1979), also known by her stage name "Bat for Lashes", is a British alternative rock musician. As a musician she sings and plays the piano, guitar, harpsichord and the autoharp.

Khan's debut album *Fur and Gold*, released on 11 September 2006, peaked at #48 on the UK Album Charts and went on to be nominated for the Mercury Prize in 2007, in the following year Khan also picked up two BRIT nominations for Best Breakthrough Artist and Best Female, leading to growing media attention.

Khan's second album *Two Suns*, released on the 6 April 2009, reached #5 on the UK Album Charts and #17 on the Irish Album Charts. The first single released off the second album "Daniel" charted within the top 40. The second single "Pearl's Dream" was released on 22 June 2009. The album *Two Suns* has also been nominated for the Mercury Prize.

[Read more at Wikipedia...](#)

Latest Tracks Played On The BBC

- Pearl's Dream**
BBC 6 MUSIC | CERY S ON 6 30/11/2009
- Daniel**
BBC RADIO 2 | ALEX LESTER 30/11/2009
- Prescilla**
BBC 6 MUSIC | SHAUN KEAVENY 26/11/2009

Audio Previews From Latest Album



Two Suns

As another example, the [BBC Wildlife Finder](#) provides a web identifier for every species (and other biological ranks), habitat and adaptation the BBC has an interest in. BBC Nature aggregates data from different sources, including Wikipedia, the [WWF's Wildfinder](#), the [IUCN's Red List of Threatened Species](#), the Zoological Society of London's [EDGE of Existence programme](#), and the [Animal Diversity Web](#). BBC Wildlife Finder repurposes that data and puts it in a BBC context, linking out to programme clips extracted from the BBC's [Natural History Unit](#) archive.

BBC Text only Help Search Explore the BBC

WILDLIFE FINDER

WATCH THE MOST AMAZING ANIMALS IN THE WORLD

LATEST WILDLIFE

Mosquitoes 20 November 2009
Mosquitoes are a family of insects encompassing over 3,500 recognised species. Most are associated with warm and humid regions, although they can be found in cooler areas and even in snow. Most mosquito species feed on nectar, however, the females of some species drink blood for the nutrients they need to develop eggs.

Beluga whale
Belugas come together in pods in arctic and sub-arctic seas. Their weight is fat.

COLLECTIONS
Collections offer a new perspective on the vast catalogue of clips showcasing presenter favourites, aspects of wildlife film-making, and gathering content together in illuminating ways.

The wildlife of Life
In autumn 2009, a major new series brought us life as we've never seen it before. Filmed in HD, the series illustrated the extraordinary lengths to which animals and plants go in order to survive.

David Attenborough's favourite moments
David Attenborough's favourite moments from the last 30 years represent not only memorable personal experiences, but also the

EDITOR'S CHOICE
This week's highlights feature the wild stars of Life and Natural World.

Sunflower star 2 videos Chimpanzee 6 videos
Giant cuttlefish 2 videos Hippopotamus 6 videos

WELCOME TO WILDLIFE FINDER
Explore a wealth of video, sounds, stories and breaking news to find out more about your favourite animals, how they live and where they live.

Wildlife Finder gives you access to an ever growing catalogue of BBC natural history programmes, with video clips from series such as: Planet Earth, Blue Planet, Life on Earth, Natural World, the Nature of Britain and many more.

WHAT'S NEW?
30 November 2009
Homepage redesign: [subscribe to the RSS feed](#) for new content alerts

Browse the image carousel to see the latest wildlife pages

To facilitate integration with the resources external to bbc.co.uk the music site reuses MusicBrainz URL slugs and Wildlife Finder Wikipedia URL slugs. This means that it is relatively straight forward to find equivalent concepts on Wikipedia/DBpedia and Wildlife Finder and, MusicBrainz and /music.

Elsewhere on the BBC website, there are still many areas that have not been created from the ground up with semantic web principles in mind. We are tagging those pages using DBpedia web identifiers. Each page is tagged with the concepts it is 'about'. Each tag is displayed as a link to an aggregation page for that concept. A news story might link to Barack Obama.

Enhancing search

The BBC Search Team is building on this newly-available linked data, by creating pan-BBC aggregations of content. These new pages are called Search+ —an indication that we are using this data to enhance the standard search experience. Each Search+ page shows the best content on its particular topic from around bbc.co.uk, and sometimes selected content from outside the BBC. The set of topics that make up the Search+ Pages essentially form a new BBC Controlled Vocabulary (CV) of concepts and entities. Each term in the BBC CV has an associated DBpedia resource, to enable us to use some of the metadata within DBpedia, and also enable links between our CV terms (and associated Search+ pages) and content both inside and outside the BBC. We also intend to use sources other than just DBpedia (e.g. Musicbrainz and Geonames) to provide these "Linked Open Data" associations in the near future.

Various groups around the BBC have contributed to the Search+ project, and we have a fledgling suite of tools to

allow people within the BBC to manage the life cycle of the Search+ pages and the data used in them. These tools allow us to: associate DBpedia resources with content pages; then promote those DBpedia resources into the BBC CV in a controlled way; build the associated Search+ pages for those terms in the BBC CV; include additional pieces of content on those Search+ pages—both from inside and outside the BBC; and to monitor the quality and usage of the Search+ pages.

Conclusions

Creating web identifiers for every item the BBC has an interest in, and considering those as aggregations of BBC content about that item, allows us to enable very rich cross-domain user journeys. This means BBC content can be discovered by users in many different ways, and content teams within the organisation have a focal point around which to organise their content.

The RDF representations of these web identifiers allow developers to use our data to build applications. The two issues, providing cross-domain navigation and machine-readable representations, are tightly interleaved. Giving access to machine-readable representations that hold links to further such representations, crossing domain boundaries, means that much richer applications can be built on top of our data, including new BBC products. In addition the system gives us a flexibility and a maintainability benefit: our web site becomes our API. Considering our feeds as an integral part of building a web site also means that they are very cheap to generate: they are just a different view of our data.

The approach has also proved to be an efficient one—allowing different development teams to concentrate on different domains while at the same time benefiting from the activities of the other teams. The small pieces loosely joined approach, which is manifest in any Linked Data project, significantly reduces the need to coordinate teams while at the same time allowing each team to benefit from the activities of others.

Key Benefits of Using Semantic Web Technology

- Usability—Making a site around the things people care and think about.
- User Experience—Having meaningful predicates and granular, addressable resources, so that those resources can be visualised in new ways.
- User Journeys—Allowing users to make their own journeys across our content. On the BBC /nature, users can start making their own documentaries. They can start on an animal, watch a programme clip, follow a link to a related habitat, read about that habitat and so on...
- One page per thing—Making our resources part of the Web and therefore linkable and discoverable.
- Our web site is our API—One URI for both machines and web browsers. Our web site can be used by third parties to create new products, e.g., [URIPlay](#), [TestTubeTelly](#), [FanHubz](#) or [Channelography](#).
- Loosely coupled development—Different teams can work together in a loosely coupled fashion. Each team focuses on their domain of interest.

© Copyright 2010, [BBC](#)